

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Characterisation of dairy strains of  
*Geobacillus stearothermophilus* and  
a genomics insight into its growth  
and survival during dairy  
manufacture**

A thesis presented in partial fulfilment of the  
requirements for the degree of  
Doctor of Philosophy  
in Microbiology  
at Massey University, Palmerston North,  
New Zealand

**Sara Burgess  
2016**

*This thesis is dedicated to my sons, Samuel and James.*

# Abstract

The thermophilic bacilli, such as *G. stearothermophilus*, are an important group of contaminants in the dairy industry. Although these bacilli are generally not pathogenic, their presence in dairy products is an indicator of poor hygiene and high numbers are unacceptable to customers. In addition, their growth may result in milk product defects caused by the production of acids or enzymes, potentially leading to off-flavours. These bacteria are able to grow in sections of dairy manufacturing plants where temperatures reach 40 – 65 °C. Furthermore, because they are spore formers, they are difficult to eliminate. In addition, they exhibit a fast growth rate and tend to readily form biofilms. Many strategies have been tested to prevent the formation of thermophilic bacilli biofilms in dairy manufacture, but with limited success. This is, in part, because little is known about the diversity of strains found in dairy manufacture, the structure of thermophilic bacilli biofilms and how these bacteria have adapted to grow in a dairy environment.

In Chapters 2 and 3, phenotypic approaches were taken to understand the diversity of strains within a manufacturing plant. Specifically in Chapter 2, strains of the most dominant thermophilic bacilli, *G. stearothermophilus*, were isolated from the surface of various locations within the evaporator section and ten strains were evaluated for different phenotypic characteristics. Biochemical profiling, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and fatty profiling demonstrated that the population was diverse. In Chapter 3, it was shown that the same ten strains varied in their ability to form biofilms and produce spores. Three strains of *G. stearothermophilus*, A1, P3 and D1, were selected for further analysis. SEM demonstrated that there were differences in biofilm morphologies between the three strains, particularly D1 versus the other two strains, A1 and P3.

In Chapters 4, 5 and 6 a comparative genomics approach was taken to determine how these bacteria are able to grow and survive within a dairy manufacturing environment, as well as how they differ from other strains of *Geobacillus*. In Chapter 4 draft genome sequences were generated for three strains of *G. stearothermophilus*. Identification of a putative lactose operon in the three dairy strains provided evidence of dairy adaptation. In Chapter 5 a phylogenomics approach was taken to resolve relationships within the *Geobacillus* genus and to identify differences within the *G. stearothermophilus* group itself. Finally in Chapter 6 comparison with the model organism *B. subtilis*, gave a genomics insight into the potential mechanisms of sporulation for *Geobacillus* spp.

# List of Publications

**Burgess S A**, Flint S H, Lindsay D, Cox M P and Biggs P J (2016). An updated analysis of *Geobacillus* taxonomy based on phylogenomic principles. Submitted to *BMC Microbiology*.

**Burgess S A**, Cox M P, Flint S H, Lindsay D and Biggs P J (2015). Draft genome sequences of three strains of *Geobacillus stearothermophilus* from a milk powder manufacturing plant. *Genome Announcements*. doi:10.1128/genomeA.00939-15.

**Burgess S A**, Lindsay D and Flint S H (2014). Biofilms of thermophilic bacilli isolated from dairy processing plants and efficacy of sanitisers. In: *Microbial Biofilms: Methods and Protocols*, Methods in Molecular Biology, vol. 1147, pp 367-377, Springer, New York.

**Burgess S A**, Flint S H and Lindsay D (2014). Characterization of thermophilic bacilli from a milk powder processing plant. *Journal of Applied Microbiology*. 11: 350-359.

**Burgess S A**, Lindsay D, Flint S H (2010). Thermophilic bacilli and their importance in dairy processing. *International Journal of Food Microbiology*. 144: 215 – 225.

# List of Presentations

**Burgess S A**, Lindsay D, Flint S H and Biggs P J (2015). Genomic insights into the adaptation of *Geobacillus stearothermophilus* to a milk powder manufacturing plant. Presented at: *New Zealand Microbiological Society Conference, Rotorua, November 2016*.

**Burgess S A**, Lindsay D, and Flint S H (2012). Biofilm formation and spore production of *Geobacillus stearothermophilus*. Presented at: *New Zealand Microbiological Society Conference, Dunedin, November 2012*.

# Acknowledgements

To my supervisors Steve Flint, Denise Lindsay, and Roberto Kolter thank-you for sticking with me over the 6 years it took to complete this journey. Steve, you always have an amazing amount of enthusiasm. Denise, thank-you for your thorough reviews and providing such a supportive environment when things got tough. Roberto, I will be forever grateful for taking me on even though you had never heard of me. Your probing questions at meetings kept me on my toes and the time I spent in your lab opened my eyes to a different world of microbiology.

To my additional supervisor Patrick Biggs, who came on board part-way through my PhD, thank-you for taking me on. Your expertise was a lifesaver and your enthusiasm gave me the boost I needed. I have learnt an enormous amount from you.

Steve Holroyd, without your support I never would have started this PhD. Thank-you for being there as a sounding-board when things got tough.

Hera Vlamakis, you are one of the most wonderful people I've met. Thank-you for making me so welcome in the Kolter lab and for not rolling your eyes when going through some basic microbial genetics.

To everyone in the Kolter lab (Erin, Pascale, Beth, Matt, Diana, Robe, Majse, Christian, Liraz, Maerite, Marta, Diego, and Ilana) thank-you for making me so welcome and providing advice when needed.

Thank-you to Fulbright New Zealand for giving me the opportunity to go to Boston through the Fulbright Science and Innovation Graduate Award.

Ashlee Earle, thank-you for getting me started with the comparative genomics.

Murray Cox, thank-you for generating the draft assemblies. I also really appreciated the thorough reviews of some of my draft manuscripts with some probing questions that got me thinking more outside the square.

Teng Yi Sin, the Massey Singapore student, thank-you for your dedication and hard-work during your project.

Doug and Jordan from the Massey University microscopy unit thank-you for your help and advice with the microscopy imaging.

Jasna Rakonjac and Wesley Wen thank-you for your advice and help with DNA extractions. Jasna I also really appreciated being able to come to your laboratory meetings and for giving advice when needed.

Anne-Marie Jackson and Julia Good I really appreciated you making me feel at ease when moving to the SEAT microbiology lab and helping me out with equipment.

Matt and Steve thank-you for the IT support.

Thank-you to the various people at the Fonterra Cooperative Group who helped out with laboratory work and sampling, including Donna Crowe, Matt Lin, Jacinda Aplin, Saloni Gokarn, Mark Reynolds, Helaina Colle, Micro-servicing staff and plant operators from Kauri and Te Rapa.

To my family I could not have done this without your love and support. To my Dad for initiating some ideas, being there when I needed a sounding board or advice. To my Mum for looking after the kids sometimes, for the dinners every Tuesday and hugs when I needed them.

To my husband Mitch, thank-you so so so much. You have been so supportive, you gave up your job so that I could spend time in the Kolter lab, you have been there for the kids at night so that I could get my sleep and you even built me a computer that made the genomic analyses so much easier. I could not have done this without you. Finally to my two sons Samuel and James, who tolerated many mornings without me.



# Significant contributions from others

- Fatty acid profiling by Saloni Gokarn and Mark Reynolds, Fonterra Research and Development Centre
- 18-h biofilm assays by Massey University Singapore student Teng Yi Sin
- Biofilm SEM imaging by Massey University microscopy unit (MMIC)
- Mi-Seq™ sequencing by New Zealand Genomics Ltd., Massey Genome Service, Massey University
- Draft genome assemblies by Murray Cox
- In-house Perl scripts and repeter analysis showing repetitive regions in the genome of *G. kaustophilus* HTA426 prepared by Patrick Biggs

# Table of Contents

Abstract .....	III
List of Publications .....	IV
List of Presentations .....	V
Acknowledgements .....	VI
Significant contributions from others .....	VIII
Table of Contents .....	IX
List of Figures .....	XIV
List of Tables .....	XVII
Non-standard abbreviations .....	XIX
Definitions .....	XXI
Description of computer programs and on-line genomic tools .....	XXIII
1 General introduction .....	1
1.1 Classification of the thermophilic bacilli .....	1
1.1.1 Taxonomy of the obligate thermophilic bacilli .....	1
1.1.2 Specific characteristics of the <i>Geobacillus</i> genus .....	2
1.1.3 Specific characteristics of <i>Anoxybacillus flavithermus</i> .....	2
1.2 Importance of thermophilic bacilli in dairy processing .....	3
1.2.1 Spoilage potential of thermophilic bacilli .....	3
1.2.2 Thermophilic bacilli as hygiene indicators .....	3
1.2.3 Controlling thermophilic bacteria in dairy processing .....	4
1.3 Endospores of thermophilic bacilli .....	5
1.3.1 Endospore structure .....	6
1.3.2 Spore formation .....	7
1.3.3 Spore resistance .....	8
1.3.4 Germination of spores .....	10
1.4 Biofilm formation .....	11
1.4.1 Attachment of spores and vegetative cells .....	12
1.4.2 Growth of attached populations .....	13

1.4.3	Spore development within biofilms.....	14
1.5	Enumeration and identification of the thermophilic bacilli in milk powder product.....	14
1.5.1	Plate count methods used to enumerate thermophilic bacilli .....	14
1.5.2	Alternative rapid methods .....	15
1.5.3	Characterization and identification of milk powder isolates.....	16
1.6	Conclusions.....	17
1.7	Objectives of this study.....	17
2	Isolation, identification and characterisation of strains of thermophilic bacilli from a milk powder manufacturing plant .....	19
2.1	Introduction.....	19
2.2	Materials and methods.....	20
2.2.1	Operation of the evaporators.....	20
2.2.2	Sampling regime .....	21
2.2.3	Identification of isolates.....	23
2.2.4	Biochemical characterisation.....	24
2.2.5	Analyses of fatty acid methyl esters (FAMES) .....	24
2.2.6	Analysis by MALDI-TOF MS.....	24
2.3	Results.....	25
2.3.1	Isolation of strains .....	25
2.3.2	Biochemical characterisation.....	28
2.3.3	Analyses of FAMES.....	30
2.3.4	MALDI-TOF MS analysis .....	32
2.4	Discussion .....	33
2.5	Conclusions.....	35
3	<i>In vitro</i> biofilm formation by <i>Geobacillus stearothermophilus</i> on stainless steel .....	38
3.1	Introduction.....	38
3.2	Materials and methods.....	40
3.2.1	Source of strains .....	40
3.2.2	Media and culture preparation.....	40
3.2.3	16-h biofilm and spore production assay .....	41
3.2.4	Plate counts of attached cells.....	41
3.2.5	Crystal violet microtitre plate assay.....	41
3.2.6	80-h biofilm and spore production assay .....	42
3.2.7	Scanning electron microscopy (SEM).....	43
3.3	Results.....	43
3.3.1	16-h biofilm and spore production assay .....	43
3.3.2	Crystal violet microtitre plate assay.....	44
3.3.3	80-h biofilm and spore production assay .....	47

3.4	Discussion .....	52
3.5	Conclusions .....	54
4	Genome sequencing of dairy strains of <i>Geobacillus stearothermophilus</i> .....	55
4.1	Introduction .....	55
4.2	Materials and methods .....	57
4.2.1	Genome sequencing of three <i>G. stearothermophilus</i> strains isolated from a milk powder plant .....	57
4.2.2	Genomic DNA extraction .....	57
4.2.3	Next Generation sequencing and assembly .....	58
4.2.4	Genome sequences .....	58
4.2.5	Gene prediction and annotation .....	60
4.2.6	Prediction of rRNA genes and generation of consensus gene sequences .....	62
4.2.7	Assigning predicted proteins to Clusters of Orthologous Groups (COGs) .....	63
4.2.8	Identification of differences in gene presence within the <i>Geobacillus</i> genus .....	63
4.2.9	Identification of genes important for lactose metabolism .....	66
4.2.10	CRISPR arrays .....	66
4.3	Results .....	67
4.3.1	Next generation sequencing and assembly outputs .....	67
4.3.2	General genomic features .....	68
4.3.3	Assembly and analysis of the rRNA genes .....	69
4.3.4	Gene prediction and annotation .....	74
4.3.5	COG for assigning protein function .....	77
4.3.6	Differentiation of the <i>G. stearothermophilus</i> dairy strains .....	79
4.3.7	Lactose metabolism .....	89
4.3.8	Differences between the three dairy strains .....	93
4.3.9	CRISPR arrays .....	94
4.4	Discussion .....	100
4.4.1	Fragmentation of genomes .....	100
4.4.2	Diversity of the rRNA operons .....	101
4.4.3	Differentiation of the <i>G. stearothermophilus</i> taxon .....	102
4.4.4	Niche adaptation of the dairy strains of <i>G. stearothermophilus</i> .....	103
4.4.5	Differences between the dairy strains .....	104
4.4.6	CRISPR arrays .....	105
4.5	Conclusions .....	107
5	Comparative analysis of <i>Geobacillus</i> taxonomy using phylogenomic tools .....	108
5.1	Introduction .....	108

5.2	Materials and methods.....	109
5.2.1	Genome sequences.....	109
5.2.2	Phylogenetic tree of 16S rRNA sequences .....	110
5.2.3	Ribosomal multi-locus sequence typing (rMLST) .....	110
5.2.4	Average nucleotide identity (ANI) and Tetra nucleotide frequency calculations.....	110
5.2.5	Core genome comparisons .....	111
5.2.6	Identification of strains based on the 16S rRNA gene.....	111
5.3	Results.....	111
5.3.1	16S rRNA gene sequencing and rMLST.....	111
5.3.2	ANI and TETRA calculations.....	115
5.3.3	Core genome comparisons .....	117
5.3.4	Classification of <i>Geobacillus</i> species .....	121
5.4	Discussion .....	122
5.5	Conclusions.....	125
6	Genomic insights into biofilm formation and sporulation of <i>Geobacillus</i> spp. ....	127
6.1	Introduction.....	127
6.2	Materials and methods.....	132
6.2.1	Genome sequences.....	132
6.2.2	Gene lists.....	135
6.2.3	Identification of homologues.....	136
6.2.4	Catergorization of histidine kinases.....	138
6.3	Results and discussion .....	138
6.3.1	Activation of the phosphorelay .....	139
6.3.2	Biofilm formation .....	146
6.3.3	Sporulation .....	156
6.3.4	Summary of results .....	185
6.3.5	Reliability of results .....	187
6.4	Conclusions.....	188
7	Conclusions and summary .....	190
7.1	Strain diversity within a dairy manufacturing plant.....	190
7.2	Taxonomic classification .....	191
7.3	Genomics insight into growth and survival in a dairy manufacturing environment.....	191
7.4	Summary .....	192
7.5	Practical contributions to the dairy industry.....	193
7.6	Future directions.....	193
8	Bibliography .....	195

Appendix 1	Preparing a BLAST database.....	235
Appendix 2	Mapping using Bowtie 2.....	237
Appendix 3	Identifying orthologous clusters .....	238
Appendix 4	Cumulative frequency of trimmed read lengths .....	239
Appendix 5	General genomic features of published <i>Geobacillus</i> strains .....	243
Appendix 6	Core genome sequence comparisons .....	244
Appendix 7	Heat-maps to visualize the presence/absence of homologues in <i>Geobacillus</i> spp. ....	245
Appendix 8	Gene organisation of the <i>skin</i> elements.....	247
Appendix 9	BLASTX analysis of the <i>rsfA</i> and <i>gerR</i> genes.....	248

# List of Figures

Figure 1.1. Structure of a <i>B. subtilis</i> spore, taken from McKenney <i>et al.</i> (2013).....	7
Figure 1.2. Schematic diagram of the sporulation process, taken from McKenney <i>et al.</i> (2013).....	7
Figure 2.1. Schematic diagram of the milk powder processing plant used for this study.....	21
Figure 2.2. Schematic diagram of one MVR effect from an evaporator.....	22
Figure 2.3. Thermophile plate counts (cfu/ml) of milk samples taken at the start (black bars) and end (grey bars) of a nutritional milk powder processing cycle.....	26
Figure 2.4. Variation in the percentage of six fatty acids across five <i>Geobacillus</i> <i>stearothermophilus</i> strains.....	32
Figure 2.5. Dendrogram based on MALDI-TOF spectra of the <i>G. stearothermophilus</i> isolates.....	33
Figure 3.1. Total thermophile counts (solid filled bars) and thermophilic spore counts (pattern filled bars) attached to the surface of a stainless steel coupon after 16 h of incubation at 55°C in RSM.....	43
Figure 3.2. SEM images of three selected <i>G. stearothermophilus</i> strains: P3 (A), D1 (B) and A1 (C); grown on the surface of stainless steel in RSM.....	44
Figure 3.3. Biofilm forming index based on the OD <sub>590nm</sub> obtained with the CV assay for the three selected strains of <i>G. stearothermophilus</i> .....	45
Figure 3.4. Planktonic cell count of three selected strains of <i>G. stearothermophilus</i> after 16- h growth in modified milk medium with the CV assay.....	46
Figure 3.5. Biofilm formation visualised by CV staining of three selected strains of <i>G.</i> <i>stearothermophilus</i> on the surface of a 6-well microtitre plate, grown in modified milk medium for 18 h.....	46
Figure 3.6. Total thermophile counts (A) and thermophilic spore counts (B) of three selected <i>G. stearothermophilus</i> strains attached to the surface of a stainless steel coupon over an incubation period of 80 h at 55°C in 1/5 diluted RSM based medium.....	48
Figure 3.7. pH readings of the medium, before it was replaced throughout an 80 h incubation period, used for growing biofilms of three selected strains of <i>G.</i> <i>stearothermophilus</i> .....	49
Figure 3.8. SEM images with a magnification of 1300x of three selected <i>G.</i> <i>stearothermophilus</i> strains: A1, P3 and D1; grown on the surface of stainless steel in 1/5 diluted RSM based medium over an 80 h incubation period at 55°C.....	51
Figure 3.9. SEM images with a magnification of 5000x of biofilms after 56 h of incubation of the three selected <i>G. stearothermophilus</i> strains: A1, P3 and D1; grown on the surface of stainless steel.....	52

Figure 3.10. SEM images with a magnification of 5000x showing spore formation within biofilms of the three selected <i>G. stearotherophilus</i> strains: A1 P3 and D1; after 8 h of incubation. ....	52
Figure 4.1 Lactose utilisation in LAB.....	56
Figure 4.2. Overall process of mode 1 from the in-house Prokka script for annotating genomes .....	62
Figure 4.3. Overall process of the in-house OrthoMCL script.....	64
Figure 4.4. Reputer analysis showing repetitive regions greater than or equal to 150bp in the genome of <i>G. kaustophilus</i> HTA426.....	68
Figure 4.5. Alignment of the 5S rRNA copies. (A) A1 and (B) D1. ....	70
Figure 4.6. Genetic context of a 5S rRNA gene from strain D1, which sits by itself, compared with the split rRNA operon found in <i>G. kaustophilus</i> HTA426.....	70
Figure 4.7. Gene organisation of a region of the genome for <i>G.stearotherophilus</i> strain P3 containing putative biofilm genes. ....	75
Figure 4.8. Gene organisation of a region of the genome for <i>G.stearotherophilus</i> strain P3 containing a CRISPR array. ....	76
Figure 4.9. Proportion of coding sequences assigned to a protein function for the three <i>G.stearotherophilus</i> dairy strains (coloured in red) compared with other strains of <i>Geobacillus</i> spp. and <i>B. subtilis</i> . ....	77
Figure 4.10. The conservation of selected genes involved in the utilisation of lactose and galactose.....	90
Figure 4.11. Comparison of the organisation of the <i>lac</i> genes. ....	92
Figure 4.12. Organisation of the CRISPR arrays and associated genes in the dairy strains A1, P3 and D1.....	97
Figure 4.13. Organisation of the spacers in each CRISPR array. ....	98
Figure 5.1. Un-rooted maximum likelihood phylogenetic trees for the <i>Geobacillus</i> genus.....	113
Figure 5.2. Heat-map comparison of 16S rRNA gene percentage identities. ....	114
Figure 5.3. Heat map comparison of the ANIm values. ....	116
Figure 5.4 Core genome sequence comparisons. ....	120
Figure 6.1. Initiation of biofilm formation and sporulation through activation of Spo0A. ....	128
Figure 6.2. Simplified schematic of the sporulation sigma factor regulatory network. ....	130
Figure 6.3. Conservation of those genes involved in the phosphorelay and their equivalent COGs in <i>Geobacillus</i> spp. ....	140
Figure 6.4. Domain organization of the orphan HKs from (A) <i>B. subtilis</i> and (B) <i>Geobacillus</i> strains A1 (Types 2, 3, 4 and 5) and NUB3621 (Type 1, 6 and 7). ....	144
Figure 6.5. Alignment of the sequence surrounding the phosphorylated histidine of orphan HKs from <i>B. subtilis</i> 168, <i>G. stearotherophilus</i> A1 and <i>Geobacillus</i> sp. NUB3621.....	144
Figure 6.6. Conservation of selected <i>B. subtilis</i> biofilm genes in <i>Geobacillus</i> spp. and their equivalent COGs. ....	148



Figure 6.7. Gene organisation of biofilm genes in <i>B. subtilis</i> and putative exopolysaccharide genes in selected <i>Geobacillus</i> strains and <i>B. cereus</i> ATCC 14579. ....	152
Figure 6.8. Gene organisation of the <i>ica</i> operon in <i>S. aureus</i> and the homologue of <i>icaA</i> in the dairy strains <i>G. stearothermophilus</i> A1 and D1.....	155
Figure 6.9. Conservation of selected <i>Bacillus subtilis</i> sporulation genes in the <i>Geobacillus</i> genus.....	158
Figure 6.10. Conservation of selected <i>B. subtilis</i> regulatory network genes in <i>Geobacillus</i> spp. and their equivalent COGs. ....	161
Figure 6.11. Gene organisation of <i>sigK</i> in <i>B. subtilis</i> 168 and selected <i>Geobacillus</i> strains. ....	164
Figure 6.12. Organisation of selected <i>B. subtilis</i> genes encoding for proteins involved with engulfment and their homologues in <i>Geobacillus</i> . ....	167
Figure 6.13. Conservation of selected <i>B. subtilis</i> cortex genes in <i>Geobacillus</i> spp. and their equivalent COGs. ....	168
Figure 6.14. Gene organisation of <i>coxA</i> (coloured in green) and <i>safA</i> (coloured in blue) in <i>B. subtilis</i> and their homologues in two strains of <i>Geobacillus</i> : <i>G. stearothermophilus</i> A1 and <i>Geobacillus</i> sp. Y4.1MC1 (Taxon group 10, <i>G. thermoglucosidasius</i> ). ....	170
Figure 6.15. Conservation of selected <i>B. subtilis</i> spore coat genes in <i>Geobacillus</i> spp. and their equivalent COGs. ....	172
Figure 6.16. Gene organisation of the <i>sps</i> operon in <i>B. subtilis</i> and putative <i>sps</i> operons in selected strains of <i>Geobacillus</i> spp. ....	176
Figure 6.17. Collagen-like glycoproteins.....	178
Figure 6.18. Gene organisation of the <i>spoIIAA-sigF</i> and <i>spoVA</i> operons of <i>B. subtilis</i> and their homologues (as depicted by the different colours) in <i>Geobacillus</i> . ....	180
Figure 6.19. Presence/absence of genes encoding for SASPs.....	183
Figure 6.20. Conservation of <i>B. subtilis</i> mother-cell lysis genes in <i>Geobacillus</i> spp. and their equivalent COGs. ....	184

# List of Tables

Table 2.1. Identification of strains isolated from the evaporator .....	27
Table 2.2. Biochemical reactions of thermophile <i>G. stearothersophilus</i> isolates based on the Microgen Bacillus-ID, BBL Crystal and API 50CHB identification kits.....	29
Table 2.3 FAME percentage composition* of the thermophile <i>G. stearothersophilus</i> isolates compared with the type strain <i>G. stearothersophilus</i> ATCC 12980.....	31
Table 3.1. Bacterial strains used in this study. ....	40
Table 3.2. Percentage of the total thermophile counts that were spores attached to the surface of a stainless steel coupon for the three selected <i>G.</i> <i>stearothersophilus</i> strains. ....	49
Table 4.1. Bacterial genomes <sup>a</sup> used in this study .....	59
Table 4.2. Options for the in-house Prokka script .....	61
Table 4.3. Description of the MySQL tables generated from the in-house Prokka script.....	62
Table 4.4. Options for the in-house OrthoMCL script.....	64
Table 4.5. Description of the MySQL tables generated by the in-house OrthoMCL script.....	65
Table 4.6. Coverage for each <i>G. stearothersophilus</i> genome sequenced .....	67
Table 4.7. Assembly statistics.....	67
Table 4.8. General features of the draft genomes for the sequenced <i>G. stearothersophilus</i> strains A1, D1 and P3.....	69
Table 4.9. Variant frequency of polymorphisms in the 5S rRNA gene sequences of the <i>G.</i> <i>stearothersophilus</i> dairy strains. ....	70
Table 4.10. Diversity of the 5S rRNA genes.....	71
Table 4.11. Variant frequency of polymorphisms in the 16S rRNA gene sequences of the <i>G.</i> <i>stearothersophilus</i> dairy strains. ....	72
Table 4.12. Variant frequency of polymorphisms in the 23S rRNA gene sequences of the <i>G.</i> <i>stearothersophilus</i> dairy strains. ....	73
Table 4.13. Diversity of the 16S rRNA genes.....	74
Table 4.14. Diversity of the 23S rRNA genes.....	74
Table 4.15. Comparison of annotation descriptions between RAST and Prokka for a region of the genome of strain P3 containing putative biofilm genes.....	75
Table 4.16. Comparison of annotation descriptions between RAST and Prokka for a region of the genome of strain P3 containing a CRISPR array.....	76
Table 4.17. Selected genes only found in <i>G. stearothersophilus</i> A1, P3 and D1.....	80
Table 4.18. Selected genes absent in <i>G. stearothersophilus</i> A1, P3 and D1.....	81
Table 4.19. COGs absent in <i>G. stearothersophilus</i> A1, P3 and D1 and present in the other <i>Geobacillus</i> spp. ....	82
Table 4.20. Selected genes absent in <i>G. stearothersophilus</i> A1, P3, D1 or ATCC 7953 but present in the other <i>Geobacillus</i> spp .....	83

Table 4.21. COGs absent in <i>G. stearothermophilus</i> and present in the other <i>Geobacillus</i> spp. ....	88
Table 4.22. BLASTX of <i>G. stearothermophilus</i> A1 <i>lac</i> cluster components.....	93
Table 4.23. CRISPR arrays found in the <i>Geobacillus</i> genus.....	94
Table 4.24. Characteristics of the CRISPR arrays.....	99
Table 5.1. Number of genes and amino acids used in the OrthoMCL clustering. ....	118
Table 5.2. Identification of <i>Geobacillus</i> strains using the RDP.....	121
Table 5.3. Identity of <i>Geobacillus</i> strains.....	122
Table 6.1. Genes involved in production of the extracellular matrix in <i>B. subtilis</i> . ....	129
Table 6.2. <i>Geobacillus</i> genomes used in this study.....	133
Table 6.3. Reference genomes used in this study. ....	134
Table 6.4. Genes encoding for exosporium proteins of <i>B. anthracis</i> .....	136
Table 6.5. Description of each COG, which contained a protein involved in the <i>B. subtilis</i> phosphorelay. ....	141
Table 6.6. Description of each COG, which contained a protein involved in biofilm formation of <i>B. subtilis</i> .....	150
Table 6.7. Putative <i>eps</i> cluster in <i>B. cereus</i> and their homologues in <i>B. subtilis</i> and presence in <i>Geobacillus</i> spp. ....	153
Table 6.8. Description of each COG, which contained a protein involved in <i>B. subtilis</i> spore regulation. ....	162
Table 6.9. Genomes containing a split <i>sigK</i> gene.....	163
Table 6.10. Description of each COG, which contained proteins involved with cortex formation. ....	169
Table 6.11. Description of each COG which contained proteins involved with <i>B. subtilis</i> coat formation. ....	173
Table 6.12. Homologues of <i>B. cereus</i> group genes encoding for exosporium proteins.....	177
Table 6.13. Description of COGs which contain proteins involved with mother-cell lysis.....	184
Table 6.14. Summary of the key components of the phosphorelay, biofilm and sporulation pathways conserved in <i>Geobacillus</i> spp.....	186

# Non-standard abbreviations

$A_w$	Water activity
ANI	Average nucleotide identity
BDBH	Bidirectional best hit
CIP	Clean-in-place
COG	Clusters of orthologous groups
DPA	Dipicolinic acid
DSI	direct steam injection
EOR	End-of-run
GFF	General file format
HK	Histidine kinase
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MCL	Markov clustering
MLVA-HRM	Multi-locus variable-number analysis - high-resolution melt analysis
MVR	Mechanical vapour recompression (evaporator)
n/d	Not determined
PHE	Plate heat exchanger
rMLST	Ribosomal multi-locus sequence typing
SFB	Static fluid bed
SOR	Start-of-run
THE	Tubular heat exchanger
TM	Transmembrane

---

T/S	Total solids
TVR	Thermal vapour recompression (evaporator)

# Definitions

Accessory genome	Additional genes that are present in some members and absent from others within a group of isolates under investigation.
Clean-in-place (CIP)	Cleaning regime after a manufacturing run.
Conditioning layer	The thin layer of proteins and exopolysaccharides that forms immediately on a surface when it is submersed in a liquid.
Core genome	A set of genes shared by all members in a group of isolates under investigation.
Direct steam injection (DSI)	A direct method of heat treatment where steam is injected into the milk.
Effect	A section of an evaporator that has the same boiling temperature.
Engulfment	Part of the sporulation process where degradation of the septal membrane (between the mother-cell and forespore), and relocation of the mother-cell membrane around the forespore occurs.
Forespore	The immature form of the spore when it is being formed within the mother cell.
Foulant	The build-up of milk proteins and calcium phosphate salts on equipment surfaces in dairy manufacturing plants.
Homologue	Genes that are descendents of the same ancestral gene but were separated by either speciation or gene duplication.
Orifice pans	Located at the top of the evaporator to distribute milk into the pass tubes.
Mother cell	The cell which houses the forespore as it matures into an endospore.
Orthologue	Genes in different species which were derived from the same ancestral gene and were separated by speciation.
Paralogue	Genes that are descendents of the same ancestral gene, but were separated by gene duplication.
Pass	A section of the effect, in the evaporator, that is made up of a set of

---

	tubes that the milk passes through.
Plate heat exchanger (PHE)	An indirect method of heat treatment that consists of a series of plates where the heating or cooling medium passes on one side, and the milk on the other.
Pseudogene	A “gene” which has lost its ability to code for a protein, generally through the accrual of mutations such as stop codons or frameshifts within the gene.
Sliding	The passive movement of bacteria across a surface. This process does not make use of bacterial appendages such as flagella or pilli.
Spore coat	The outer layers of the endospore.
Spore cortex	The layer between the inner and outer membranes of the spore and is composed of peptidoglycan.
Spore crust	The outer layer of the coat in spores of <i>Bacillus subtilis</i> .
Spore exosporium	The outermost layer of spores in some species of <i>Bacillus</i> . It is composed of glycoprotein and separated from the coat by a large irregular space.
Swarming	The coordinated movement, through the use of flagella, of a bacterial population across a surface.
Water activity	In the dairy context this refers to the amount of water not bound to food molecules. This water can enable the growth of bacteria. When milk powder is made the water activity decreases through the evaporators as the milk is concentrated and once dried reaches levels that no longer supports bacterial growth.

# Description of computer programs and on-line genomic tools

Bowtie 2	An alignment program used for aligning short sequences (e.g. sequence reads from a genome sequencer) to long sequences (e.g. genome sequences) (Langmead & Salzberg, 2012). The output generated by Bowtie 2 is a SAM file.
COGnitor	A software tool designed to assign predicted proteins to the already established COGs (Tatusov <i>et al.</i> , 2000, Galperin <i>et al.</i> , 2015).
CRISPRDetect	An on-line tool ( <a href="http://brownlabtools.otago.ac.nz/CRISPRDetect/predict_crispr_array.html">http://brownlabtools.otago.ac.nz/CRISPRDetect/predict_crispr_array.html</a> ), designed to detect the presence of CRISPR arrays (Biswas <i>et al.</i> , 2014).
CRISPRTarget	An on-line tool ( <a href="http://brownlabtools.otago.ac.nz/CRISPR_WEB/crispr_analysis.html">http://brownlabtools.otago.ac.nz/CRISPR_WEB/crispr_analysis.html</a> ), designed to determine the target of CRISPR spacers (Biswas <i>et al.</i> , 2013).
GET_HOMOLOGUES	A software package that incorporates three different algorithms (BDBH, COGtriangles and OrthoMCL) for clustering homologous genes (Contreras-Moreira & Vinuesa, 2013).
Jspecies	A software package designed for comparing the similarity of two or more bacterial species (Richter & Rossello-Mora, 2009). Synthetic DNA-DNA hybridisations can be carried out using three methods: Average nucleotide identity (ANI) calculated using BLAST, ANI calculated using MUMmer and calculation of tetra nucleotide frequencies (TETRA).
OrthoMCL	A software program which uses an algorithm incorporating both BLASTP and the Markov clustering algorithm to determine orthologous groups of proteins within a group of genomes (Li <i>et al.</i> , 2003).
Pfam	Pfam ( <a href="http://pfam.xfam.org/">http://pfam.xfam.org/</a> ) is a database of protein families. In this present study it was used for identifying domains in predicted protein sequences.
Prokka	Prokka is a software package used for rapidly annotating prokaryotic



genomes (Seemann, 2014).

Rapid Annotation  
using Subsystem  
Technology (RAST)

A web based server (<http://rast.nmpdr.org/>), which can carry out automated annotations on bacterial genomes (Aziz *et al.*, 2008).

RNAmmer

A web based server (<http://www.cbs.dtu.dk/services/RNAmmer/>), used to predict prokaryotic and eukaryotic rRNA gene sequences in genome sequences (Lagesen *et al.*, 2007).

Velvet

An algorithm package used for *de novo* genome assembly (Zerbino & Birney, 2008). In assembling, the sequence reads are broken into shorter sequences called k-mers and used to generate de Bruijn graphs. A range of k-mer lengths are tested to generate the best assembly.